International Journal of Wireless Information Networks, Vol. 9, No. 3, July 2002 (© 2002)

A Probabilistic Approach to WLAN User Location Estimation

Teemu Roos,[1,3] Petri Myllymäki,[1] Henry Tirri,[1] Pauli Misikangas,[2] and Juha Sievänen[2]

We estimate the location of a WLAN user based on radio signal strength measurements performed by the user's mobile terminal. In our approach the physical properties of the signal propagation are not taken into account directly. Instead the location estimation is regarded as a machine learning problem in which the task is to model how the signal strengths are distributed in different geographical areas based on a sample of measurements collected at several known locations. We present a probabilistic framework for solving the location estimation problem. In the empirical part of the paper we demonstrate the feasibility of this approach by reporting results of field tests in which a probabilistic location estimation method is validated in a real-world indoor environment.

KEY WORDS: Location estimation; mobile terminals; WLAN; machine learning; probabilistic modeling.

1. INTRODUCTION

Location-aware computing is a recent interesting research area that exploits the possibilities of modern communication technology [1–4]. Location-aware devices can be located or can locate themselves; by location-aware services we mean services based upon such location technologies. Location-aware computing has great potential in areas such personal security, navigation, tourism, and entertainment. The most obvious location-based service is the one answering questions like "Where am I?" and "Where is the nearest shop/bus-stop/fire-exit?". Using graphical and interactive terminals it is possible to implement an application presenting a map labeled with a mark pointing "You are here". Furthermore, location information can also be useful for other people than the user of the location-aware device. It is often useful to be able to locate a group of other people, e.g. friends, co-workers, or customers.

The location of a mobile terminal can be estimated using radio signals transmitted or received by the terminal [5–11]. The problem has various names: location estimation, geolocation, location identification, location determination, localization, positioning, and so on. Some location estimation methods, such as GPS, are based on signals transmitted from satellites, whereas others rely on terrestrial communication. Additional costs to the service provider are minimal in systems based on existing network infrastructure. For instance, the cell-ID method, in which the location of the nearest base station is reported as a location estimate, is applicable in most networks. However, the location estimation accuracy of such systems is often inadequate for many location services. Improving the accuracy of location estimation systems based on the existing network infrastructure would be very useful, and it is the main motivation of this work. We focus primarily on wireless local area networks (WLANs), but most of the ideas and concepts are applica-

[1] Complex Systems Computation Group, Helsinki Institute for Information Technology, P. O. Box 9800, 02015 HUT, Finland. http://cosco.hiit.fi

[2] Ekahau Inc., Salomonkatu 17 B, 00100 Helsinki, Finland. http://www.ekahau.com

[3] Phone: +358-9-85011286. Fax: +358-9-6949768

1068-9605/02/0700-0155/0 © 2002 Plenum Publishing Corporation

ble to many other networks as well, including those based on GSM/GPRS, CDMA, or UMTS standards.

The traditional, geometric approach to location estimation is based on angle and distance estimates from which a location estimate is deduced using standard geometry. We will discuss location estimation from a point of view that is different from the traditional one. Our probabilistic approach is based on an empirical model that describes the distribution of received signal power at various locations. The model is used to estimate the mobile unit's location when the received power is observed. The use of probabilistic models provides a natural way to handle uncertainty and errors in signal power measurements. Our approach is very similar to that used in [8], but we address the location estimation problem in a more general setting whereas Castro et al. focus on the problem of identifying the room where the user is in. We also demonstrate the feasibility of our approach in a systematic empirical case study in which an average location estimation accuracy of less than 2 meters was achieved.

The paper is organized as follows: We shall first explain the basic principles of the probabilistic approach in Section 2; more discussion on the probabilistic approaches to density estimation and predictive modeling in general can be found in [12–15]. In Sections 3–5 we describe some location estimation methods based on the approach. In Section 6 we present a case study in which the methods are applied in a real-world indoor test environment. The conclusions are summarized in Section 7.

2. LOCATION ESTIMATION AS A MACHINE LEARNING PROBLEM

Machine learning can be characterized as the task of automatic learning from examples. In location estimation, machine learning can be used in the following way. We first collect a set of *calibration data* consisting of signal measurements collected from various locations, each measurement labeled with the correct location. The calibration data is used in constructing a *model*, which can be later used as an estimator of the unknown location given some new signal measurements. In machine learning terms, such a procedure is often called *pattern recognition* or *pattern classification*. For a classic text on pattern recognition see [16].

In the so-called *testing phase*, location estimation performance is measured using some *loss* (or error) *function* based on the location estimate and the true location. Testing is based on a set of *test data* collected independently of the calibration data. In the location estimation domain, natural loss functions are obtained from the distance between the location estimate and the true location and its positive powers, in particular the square of the error.

Various machine learning methods can be applied in the location estimation domain. In *case-based* methods, for instance, the training examples or a part thereof are stored in a database that is accessed during the location estimation process. A prime example of a case-based method is the nearest neighbor method, which we will discuss in Section 3.

To describe an alternative, probabilistic approach, we will now introduce some notation. We denote random variables and their values by the same lowercase letters. In particular, l denotes location, and o denotes an observation variable or vector. We assume that the observation variable is a vector of received signal strength values for a set of access points in a communication network. The training data D consists of n examples, denoted by $(l_i\ o_i)$, for $i \in \{1, \ldots, n\}$, where n is the number of training examples. With a slight abuse of notation, we use the general notation $p(\cdot)$ to denote all probability distributions, for either discrete or continuous variables. Conditional probabilities are denoted by $p(\cdot|\cdot)$.

In this work we are mainly interested in the use of probabilistic models for the location estimation problem. In particular, we use models that estimate the probability distribution of the observation variable given the value of the location variable. In other words, for any given location l we can obtain a distribution $p(o|l)$. By application of the Bayes rule, we can then obtain the so-called *posterior distribution* of the location:

$$p(l|o) = \frac{p(o|l)\,p(l)}{p(o)} = \frac{p(o|l)\,p(l)}{\sum_{l' \in \mathcal{L}} p(o|l')\,p(l')} \qquad (1)$$

where $p(l)$ is the *prior probability* of being at location l before knowing the value of the observation variable, and the summation goes over the set of possible location values, denoted by $\mathcal{L}$. If the location variable is continuous, the sum should be replaced by the corresponding integral. The prior distribution $p(l)$ gives a principled way to incorporate background information such as personal user profiles and to implement tracking. For simplicity we use here only uniform priors that introduce no bias toward any particular location. Because the denominator $p(o)$ does not depend on the location variable l, it can be treated as a normalizing constant whenever only relative probabilities or probability ratios are required.

The term $p(o|l)$ is called the *likelihood function* because it gives the probability of the observation given the assumed source of the observation, in our case the location. There are several implementation possibilities

for the estimation of the likelihood function from data. In Sections 4 and 5 we present two examples, the kernel method and the histogram method. The prior being uniform, the likelihood function completely determines the posterior distribution of the location. Therefore it is of utmost importance to obtain a likelihood function that describes the distribution of the observables at all locations as well as possible.

In principle it is also possible to obtain a likelihood function without any calibration data by using knowledge of radiowave propagation and the environment. Several propagation prediction or *cell planning* tools are available for the purpose [17–20]. A few experiments of location estimation based on propagation prediction have been presented [5,21]. The results correspond to our experiences—which we will not elaborate in the present paper—suggesting that the propagation prediction–based methods are competitive against the traditional, geometric methods (see below) but not against the machine learning approach.

The posterior distribution $p(l|o)$ can be used to choose an optimal estimator of the location based on whatever loss function is considered to express the desired behavior. For instance, the squared error penalizes large errors more than small ones, which is often useful. If the squared error is used, the estimator minimizing the expected loss is the expected value of the location variable

$$\mathbb{E}[l|o] = \sum_{l' \in \mathcal{L}} l' \, p(l'|o) \tag{2}$$

assuming that the expectation of the location variable is well defined, i.e., the location variable is numerical.

The presented probabilistic approach can be contrasted with the more traditional, geometric approach to location estimation used in methods such as angle-of-arrival (AOA), time-of-arrival (TOA), and time-difference-of-arrival (TDOA). In the geometric approach the signal measurements are transformed into angle and distance estimates, from which a location estimate is deduced using standard geometry. To obtain the angle and distance estimates, one needs implicitly to have a model describing the dependency between the location and the observables, which in our probabilistic setting corresponds to the likelihood function. One of the drawbacks of the geometric approach is that there is no principled way to deal with the incompatibility of the angle and distance estimates caused by measurement errors and noise. On the other hand, the geometric approach is usually computationally very efficient. Nevertheless, in Section 5 we present a probabilistic location estimation method that is sufficiently efficient for virtually all practical purposes.

3. NEAREST NEIGHBOR METHOD

For comparison purposes we will now present the so-called *nearest neighbor method*. It is based on some context-dependent distance measure that assigns a non-negative distance value between any two observation vectors. We will use the simple Euclidean distance evaluated from observation vectors, i.e., the received signal strength values of various access points. A special heuristic is required for handling the missing values associated with the cases in which the signal of some access points are not observed at all. In this work we chose to simply replace the missing values with some constant smaller than any of the measured values. Given a set of training data and a test observation vector, the location estimate is obtained from the training example whose observation vector has the minimal distance when compared with the test observation; hence the name "nearest neighbor".

The nearest neighbor method has been used for location estimation [5,22,23]. Bahl et al. pre-process the training data by combining all examples collected from the same location into one training example whose observation vector is the mean vector of the combined vectors. The pre-processing enables faster location estimation and presumably reduces the effect of random fluctuations in the training data.

4. KERNEL METHOD

In the *kernel method* a probability mass is assigned to a "kernel" around each of the observations in the training data. Thus the resulting density estimate for an observation o in location l is a mixture of n_l equally weighted density functions, where n_l is the number of training vectors in l:

$$p(o|l) = \frac{1}{n_l} \sum_{l_i = l} K(o; o_i) \tag{3}$$

where $K(\cdot; o_i)$ denotes the *kernel function*. One widely used kernel function is the Gaussian kernel

$$K_{Gauss}(o; o_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(o - o_i)^2}{2\sigma^2}\right) \tag{4}$$

where σ is an adjustable parameter that determines the width of the kernel. Figure 1 illustrates the effect of the parameter σ.

In our location estimation domain, the density estimates are constructed for the received signal strength value. As in the nearest neighbor method, we replace the missing values by a small constant. The above one-

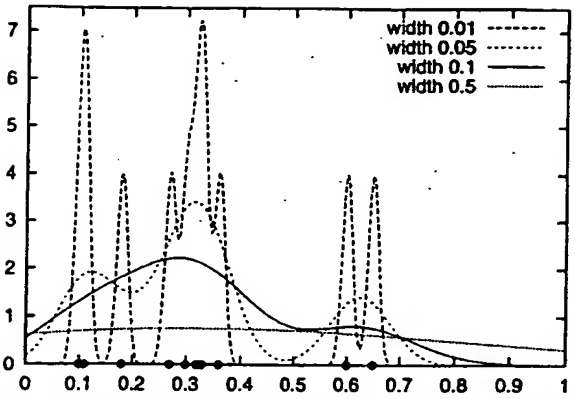


Fig. 1. Examples of kernel density estimates with Gaussian kernel and different values of the kernel width σ. The larger the value of σ, the smoother the estimate. The observed values are (0.1, 0.11, 0.18, 0.27, 0.3, 0.32, 0.33, 0.36, 0.6, 0.65).

dimensional formulas can be extended to multivariate observations, e.g., received power from several access points, by multiplying the individual probabilities, which amounts for an assumption of independence of the observations. Although the independence assumption can be criticized, it is significantly easier to estimate one-dimensional densities than multivariate densities. Moreover, the independence is only assumed locally, i.e., given the location, not globally. In other words, the components of the observation vector can, and usually do, have dependencies if the value of the location variable is not fixed.

In the kernel method the training examples can be dealt with in two ways. First, we can group the examples in clusters, each taken to be collected from a single location. Alternatively, all the examples can be considered as being collected from different locations. In the latter case, all the received signal power density estimates are based on a single observation. In our experiments the latter kernel method produced better results—all the results reported in Section 6 were obtained by using this type of individual kernels.

It is interesting to note that the Euclidean nearest neighbor method is obtained as a limiting case from the kernel method with the Gaussian kernel as the kernel width σ approaches zero. This can be seen by observing that the probability $p(l|o)$ is proportional to the likelihood $p(o|l)$, which is a Gaussian density function. Thus the probability $p(l|o)$ is a monotonically decreasing function of the squared distance between the observed signal power and the kernel mean. As the inverse of the kernel width σ grows, the squared error is multiplied by a larger value and the difference between the most probable location and the other locations grows exponentially.

5. HISTOGRAM METHOD

The so called *histogram method* is another method for estimating density functions. Its use for location estimation has been independently suggested in [7,8,11]. The histogram method is closely related to *discretization* of continuous values to discrete ones. Let us first assume that the observation variable is one-dimensional, and that the minimum and maximum of the variable are known. The method requires that we fix a set of *bins*, i.e., a set of non-overlapping intervals that cover the whole range of the variable from the minimum to the maximum. The number of the bins, denoted by k, is an adjustable parameter. The density estimate is then a piecewise constant function where the density is constant within each of the bins.

In addition to the number of bins, it is obviously necessary to fix the boundary points of the bins—a choice that greatly affects the resulting density estimate. For simplicity, here we use equal-width bins $[min + iw, min + (i + 1)w]$, $0 \leq i < k$, where min is the minimum of the observation values, and w is given by $(max - min)/k$, where max is the maximum of the observation variable. Within these constraints a histogram density is uniquely described by k parameters defining the *bin probabilities*, i.e., the value of the density function within each of the bins.

There are several alternative ways to determine good bin probabilities based on a set of observed data. In the so-called maximum likelihood method, which is probably the simplest of them all, the relative frequencies of the bins are used as the bin probabilities. A Bayesian solution (for which there are elaborate theoretical justifications, see e.g., [13]) is to add a small fraction of the total probability mass uniformly to all bins. An often reasonable fraction is given by $1/n$, where n is the size of the observed data. Such an initial probability in all bins prevents the sometimes problematic zero probabilities that are possible in the maximum likelihood method. Figure 2 presents examples of histogram densities with parameters chosen using the Bayesian solution.

Using a k bin histogram is in effect equivalent to discretization into k distinct values, each of which is assigned a point mass. The difference between values of density functions versus probability mass functions disappears as a proportionality factor. The missing values can be treated simply as the $(k + 1)$th value whose probability is estimated along with the non-missing values.

6. EMPIRICAL RESULTS

To empirically compare the location estimation methods, the nearest neighbor, kernel, and histogram

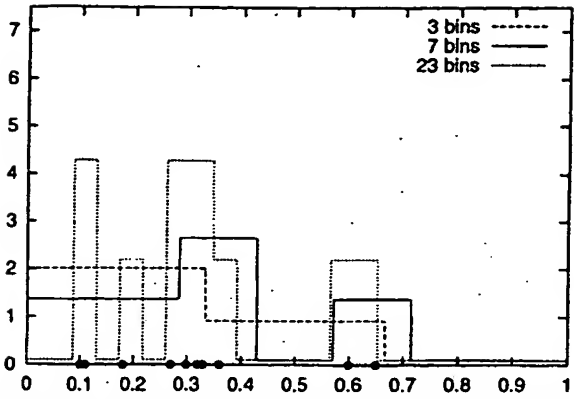


Fig. 2. Examples of histogram density estimates with different numbers of bins. The observed values (0.1, 0.11, 0.18, 0.27, 0.3, 0.32, 0.33, 0.6, 0.65) are the same as in Fig. 1.

methods were implemented as described above. We emphasize that all the adjustable parameters, such as kernel width and number of bins, were permanently fixed *before* running any tests or looking at the test data based only on calibration data. Adjusting the parameters based on test data and/or test results will usually result in overly optimistic results. The relative location estimation accuracies of the methods were assessed in the following case study. The test area consisted of a typical one-floor office (16 × 40 meters) with concrete, wood, and glass structures, and normal environmental conditions varying with the number of people in the office and their location, air humidity, temperature, etc. There were 10 access points from two different vendors. Six of them had two omnidirectional antennas, and the other four had one directional antenna (Fig. 3).

A fair comparison of the performance of different methods is difficult because there are no standardized test procedures in this domain. The empirical results are affected by decisions such as whether the located terminal is stationary or moving at a certain speed; whether the location estimation method keeps track of the location and exploits measurement history, not just the current measurement; whether the true location of the terminal is restricted to points from which calibration data is collected; whether one or several measurements are used; and many more. For instance, Bahl et al. [5] acknowledge this problem and report several different accuracies depending on the exact method by which the accuracy is measured. The experimental setup described below can be seen as a step toward defining a framework that could be used for comparing empirical results obtained by different researchers.

To eliminate the effect of randomness of human behavior, in this study the training data was collected systematically by using a 2-meter grid, and at each grid point, which we call *calibration points*, 20 observations were recorded, each consisting of received signal power values for all observed base stations. This was done twice, 5 days in between, resulting in a training data set containing 155 calibration points, 40 observations in each. The data gathering was performed by using a standard laptop computer with a WLAN card, and the process took approximately 2 hours. The test data were collected independently on the latter day with the same hardware by using a similar 2-meter grid, but by selecting the test points to be as far as possible from the calibration points, i.e., to be in the middle of the training grid. At each of the 120 test points, 20 observations were gathered.

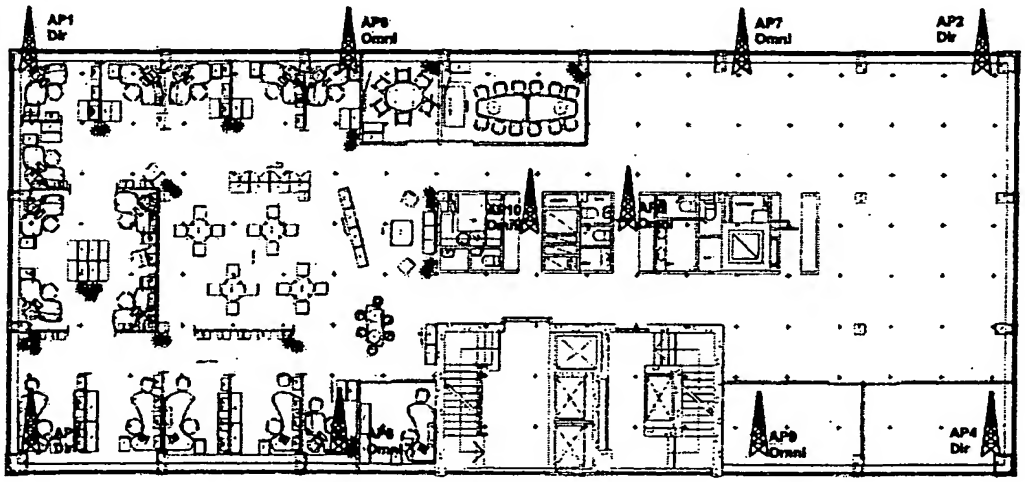


Fig. 3. The test area used in the experiments.

In the test phase, at each test point the location produced by the tested positioning method was first computed and then compared to the correct coordinates. The error was measured by using the Euclidean distance. The observation history was taken into account so that at each point, after having observed n test observations, the point estimate was smoothed to be the average of the corresponding n location estimates. More elaborate tracking schemes for handling the observation history are of course possible, but in this study this simple procedure was adopted in order to guarantee fairness in the comparison among the three different location estimation methods.

In Fig. 4a we see how the average error (averaged over all the test points) behaves as a function of the length of the history. If the time difference between two observations is, say, 100 milliseconds, we see that in 2 seconds (by using a history of 20 observations) the error drops to approximately 1.5 meters from the initial 3–4 meters obtained without history. With a short history (fast moving objects), the probabilistic methods were more accurate than the nearest neighbor method, while with the full history with 20 observations (slowly moving objects), the accuracy was approximately the same (see also Table I).

Figure 5a plots the 90 percentile error, which means that 90% of the test cases fall under this curve. The results are similar to the average results, which means that the methods are reliable in the sense that the variance of the location accuracy is relatively small.

To see how the results change with the number of access points used, we ran a series of experiments in which the data from 1–9 access points were excluded

Table I Location Estimation Errors (average, 50th, 67th, and 90th percentiles in meters) Obtained with the Nearest Neighbor, Kernel, and Histogram Methods Using 1 or 20 Test Observations (the boldface values indicate the best accuracy in each setting).

1 Test Observation				
Method	Average	50%	67%	90%
Nearest neighbor	3.71	3.21	4.38	7.23
Kernel method	**2.57**	**2.28**	**2.97**	**4.60**
Histogram method	2.76	2.32	3.11	5.37
20 Test Observations				
Method	Average	50%	67%	90%
Nearest neighbor	1.67	1.60	2.04	2.80
Kernel method	1.69	1.56	2.01	3.07
Histogram method	**1.56**	**1.45**	**1.81**	**2.76**

both in the training data and in the test data. The location accuracy was found to be surprisingly robust with respect to the number of access points used: as an illustrative example, consider Figs. 4b and 5b in which the average and 90% errors, respectively, are plotted as a function of the length of the history when only three access points were used. The three access points used in the experiment corresponding to these figures are the three access points that produced the best results, but in the exhaustive tests performed it was observed that the selection of the access points is not critical as long as the access points are not located very close to each other, in which case three base stations would not be enough to cover the whole test area.

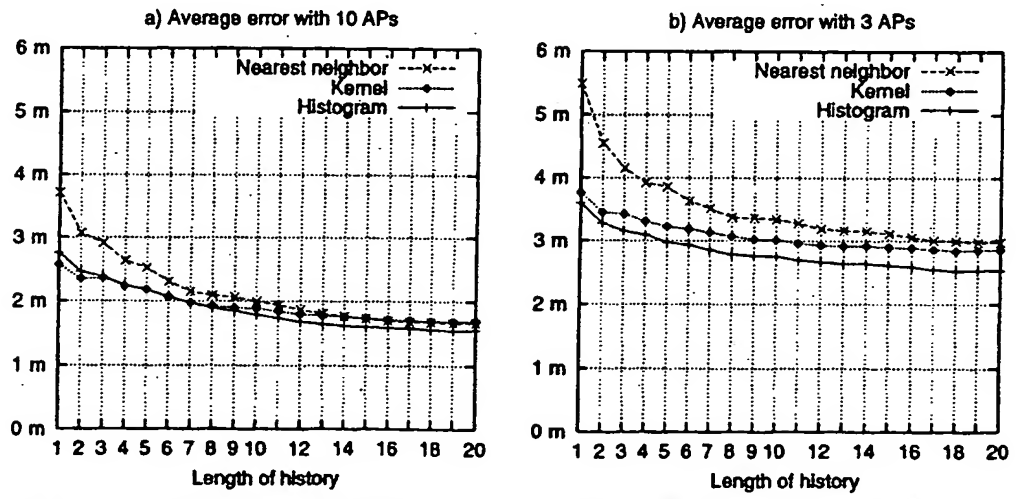


Fig. 4. Average location estimation error obtained with the nearest neighbor, kernel, and histogram methods as a function of the length of the history, measured with 10 active access points (a) and with 3 access points (b).

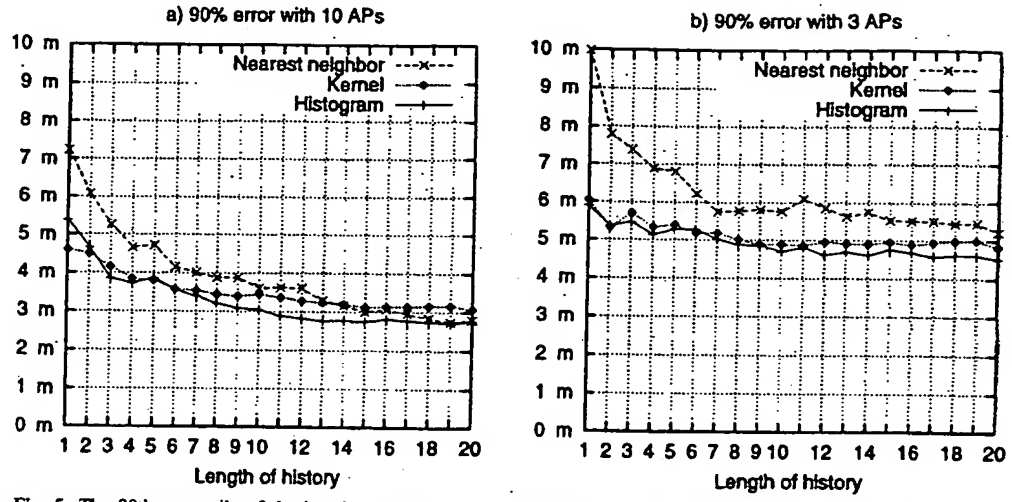


Fig. 5. The 90th percentile of the location estimation error obtained with the nearest neighbor, kernel, and histogram methods as a function of the length of the history, measured with 10 active access points (a) and with 3 access points (b).

In summary, all three methods performed well in our experiments, with the probabilistic histogram method leading to slightly lower location error on the average, especially when the number of access points was low. To examine the feasibility of the machine learning approach further, in the following we study the robustness of this method in more detail. Namely, for practical applications the optimal obtainable accuracy is often not the most important goal, but the issue is how easy it is to obtain a practically applicable accuracy. The contour plot in Fig. 6 attempts to answer this question by demonstrating the average error as a function of the number of access points and the length of the history. From this plot we see, for example, that if the full 20-observation history is available, only 7 access points are needed for obtaining an average accuracy below 2 meters, and if all 10 access points are active, only one third of the history (seven observations) are required for this level accuracy. If, on the other hand, no history is available, with 7 access points the error is increased almost to the 3-meter level, and if in this case only 5 access points would be active, the error would increase over the 3-meter level.

As mentioned earlier, the calibration points were placed systematically on a 2-meter grid to eliminate the effect of random human behavior in the data gathering process. However, although this type of data collecting process may be acceptable for scientific empirical comparisons, for real-world situations it may impractical. To simulate more realistic data gathering processes we ran a series of experiments in which a portion of calibration points were excluded in the experiments. The excluded calibration points were chosen randomly, but also at the same time in such a way that the remaining points would be distributed relatively evenly in the area. The result corresponds roughly to normal human behavior when given the data gathering task.

When all the 155 calibration points were used, the area covered by a single calibration point was on the average a little bit below 5 square meters. In Fig. 7 we see how the average error behaves when a portion of

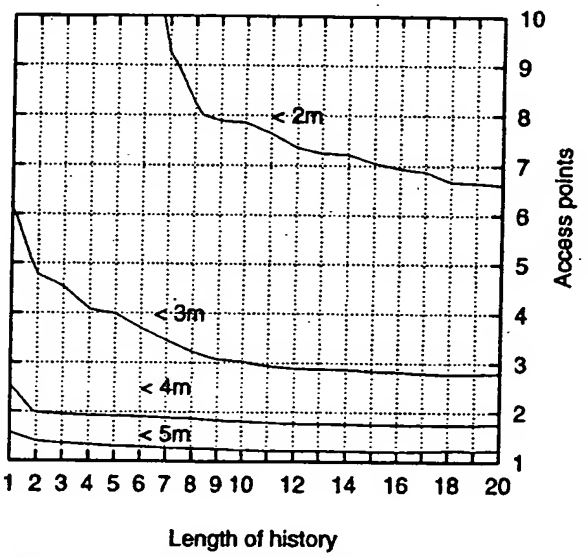


Fig. 6. Average error of the histogram method using different number of access points and test observations. The curves indicate areas corresponding to setting combinations for which the average error remains below 2, 3, 4 and 5 meters.

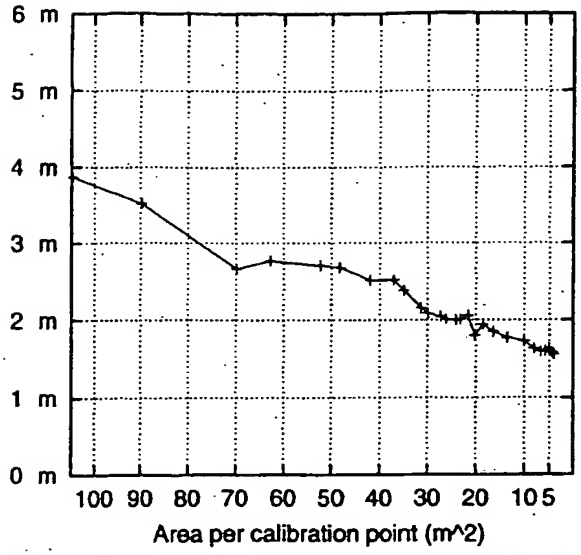


Fig. 7. Average error of the histogram method with 10 access points as a function of the density of the calibration points.

the calibration points were excluded as described above, meaning that the area covered by each point was increased. From this plot we can observe that, for example, in order to keep the average error below 2 meters, the average area covered by a single calibration point can be up to 25 square meters. If the calibration points are further pruned so that the average coverage area grows to 80 square meters, the average error remains still below three meters.

7. CONCLUSIONS

We studied the WLAN user location estimation problem in the machine learning framework in which the physical properties of the wireless networks are not considered directly, but the problem is solved by using an inductive inference procedure based on a set of training data, a database of signal measurements in known locations. Three different machine learning approaches were considered in this paper: the non-probabilistic nearest neighbor method and two probabilistic approaches. In the empirical part of the paper we compared the performance of the three different methods in a series of experimental tests. The results show that this type of a machine learning approach is feasible in the sense that a moderate amount of effort used in collecting training data produces practically applicable results—an average location estimation error below 2 meters is easy to obtain. In the experiments performed, the two probabilistic methods produced slightly better results than the nearest neighbor method.

The probabilistic methods were found to be relatively robust with respect to the number of base stations used, the amount of calibration data collected, and the length of the history used in the location estimation. On the other hand, it should be acknowledged that in real-world environments there are several environmental factors that change over time, which may cause the estimation accuracy to decrease so that recalibration is needed from time to time. Nevertheless, our initial field tests indicate that the suggested methods are relatively robust with respect to naturally occurring fluctuations, so that practical applications based on these location estimation methods are quite feasible.

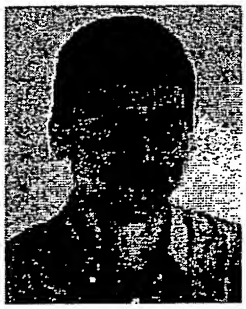
ACKNOWLEDGMENTS

This research has been supported by the National Technology Agency and the Academy of Finland.

REFERENCES

1. P. J. Brown, J. D. Bovey, and X. Chen, "Context-Aware Applications: From the Laboratory to the Marketplace," *IEEE Pers. Commun.*, Vol. 4, pp. 58–64, October 1997.
2. G. Chen and D. Kotz, "A Survey of Context-Aware Mobile Computing Research," Technical Report TR2000-381, Department of Computer Science, Dartmouth College, November 2000.
3. J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing," *Computer*, Vol. 34, No. 8, (Special Issue on Location-aware Computing pp. 57–66, August 2001.
4. U. Leonhardt, "Supporting Location-Awareness in Open Distributed Systems," Ph.D. Thesis, Department of Computing, Imperial College, London, May 1998.
5. P. Bahl and V. N. Padmanabhan, "Radar: An In-Building RF-Based User Location and Tracking System," *Proc. IEEE INFOCOM 2000*, Vol. 2, pp. 775–784, Tel-Aviv, Israel, March 2000.
6. N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less Low Cost Outdoor Localization for Very Small Devices," *IEEE Pers. Commun.*, Vol. 7, pp. 28–34, October 2000.
7. P. Myllymäki, T. Roos, H. Tirri, P. Misikangas, and J. Sievänen, "A Probabilistic Approach to WLAN User Location Estimation," *Proceedings of the 3rd IEEE Workshop on Wireless Local Areas Networks*, Boston, USA, September 2001.
8. P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A Probabilistic Room Location Service for Wireless Networked Environments," *3rd International Conference on Ubiquitous Computing (Ubicomp 2001)*, Atlanta, September–October 2001.
9. T. S. Rappaport, J. H. Reed, and B. D. Woerner, "Position Location Using Wireless Communications on Highways of the Future," *IEEE Commun. Mag.*, Vol. 34, pp. 33–41, October 1996.
10. J. Syrjärinne, "Studies of Modern Techniques for Personal Positioning," D.Sc. Thesis, Tampere University of Technology, Tampere, 2001.
11. M. A. Youssef, A. Agrawala, A. U. Shankar, and S. H. Noh, "A Probabilistic Clustering-Based Indoor Location Determination System," Technical Report CS-TR-4350 and UMIACS-TR-2002-30, University of Maryland, March 2002.

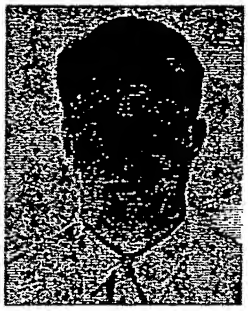
12. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.
13. J. M. Bernardo and A. F. M Smith; *Bayesian theory*. John Wiley & Sons, New York, 1994.
14. D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons, New York, 1992.
15. P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri, and P. Grünwald, "On predictive distributions and Bayesian networks," *Stat. and Comput.*, Vol. 10, pp. 39–54, 2000.
16. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley Interscience, New York, 1973.
17. J. B. Andersen, T. S. Rappaport, and S. Yoshida, "Propagation Measurements and Models for Wireless Communications Channels," *IEEE Commun. Mag.*, Vol. 33, pp. 42–49, January 1995.
18. B. H. Fleury and P. E. Leuthold, "Radiowave Propagation in Mobile Communications: An Overview of European Research," *IEEE Commun. Mag.*, Vol. 34, pp. 70–81, February 1996.
19. R. Pattuelli and V. Zingarelli, "Precision of the Estimation of Area Coverage by Planning Tools in Cellular Systems," *IEEE Personal Commun.*, Vol. 7, pp. 50–53, June 2000.
20. G. Wölfle and F. M. Landstorfer, "Prediction of the Field Strength inside Buildings with Empirical, Neural, and Ray-optical Prediction Models," *7th COST-259 MCM-Meeting*, Thessaloniki, Greece, 1999.
21. T. Tonteri, "A Statistical Modeling Approach to Location Estimation," Master's thesis, Department of Computer Science, University of Helsinki, May 2001.
22. P. Bahl, V. N. Padmanabhan, and A. Balachandran, "Enhancements to the RADAR User Location and Tracking System," Technical Report MSR-TR-00-12, Microsoft Research, February 2000.
23. K. Pahlavan, X. Li, and J.-P. Mäkelä, "Indoor Geolocation Science and Technology," *IEEE Commun. Mag.*, Vol. 40, No. 2, pp. 112–118, 2002.



Teemu T. Roos (former Tonteri) received the M.Sc. degree from the University of Helsinki in 2001. He is currently pursuing a Ph.D. degree in Computer Science. In 1999 he spent a period of 3 months at Centro Studi e Laboratori Telecomunicazioni in Turin, Italy. Since 1999 he has been with the Complex Systems Computation Group (Helsinki Institute for Information Technology). His current research interests are primarily in Bayesian and information-theoretic data analysis, and machine learning.

Petri J. Myllymäki obtained his M.Sc. degree in Computer Science at University of Helsinki in 1991 and his Ph.D. in 1995. Dr. Myllymäki is one of the cofounders of the Complex Systems Computation (CoSCo) research group, and his current special research interests are in Bayesian and information-theoretic modeling, in particular with models such as Bayesian networks or finite mixture models, and in stochastic optimization methods. He has been an editorial board member, program committee member, and reviewer for several international scientific journals and conferences, and he has published over 50 scientific articles in his research area. He has also worked as a project manager in numerous applied research projects with companies such as Nokia, Kone, StoraEnso, ABB, and AlmaMedia, and the cooperation has led to fielded applications and patent applications. Dr. Myllymäki is currently working as a Research Fellow for the Academy of Finland.

Henry R. Tirri received his B.Sc., M.Sc., and Ph.D. in Computer Science from University of Helsinki. Dr. Tirri's academic experience includes both research and teaching positions at the University of Helsinki, University of Texas at Austin, Microelectronics and Computer Technology Corporation (MCC), AT&T Bell Laboratories, Purdue University, NASA Ames Research Center, and Stanford University, where he is currently a visiting professor for 2001–2002. From 1998 he has been a professor of computer science at the University of Helsinki and an adjunct professor of computational engineering at the Helsinki University of Technology. He is currently engaged in research on various aspects of learning and adaptation in artificial systems. In particular he is interested in methods for building predictive models from data using Bayesian and information-theoretic approach. Dr. Tirri was a member of the Editorial Board of the Computer Journal (Oxford University Press) for the period 1995–1998. He was a member of the IT-development Board of the Safety Technology Authority for the period 2000–2001, and is currently a member of the Ministry of Education Board for Learning Environments as part of the National Information Technology Strategy in Education 2000–2004 and a member of the Scientific Advisory Board of WSOY Publishing Corporation. Dr. Tirri is a member of various professional societies such as ACM, IEEE and IEEE Computer Society, International Neural Network Society, AAAI, International Society for Bayesian Analysis, and American Educational Research Association.



Pauli Misikangas received the M.Sc. degree from the University of Helsinki in 1998. Since 1994 he has been developing a shogi program called 'Shocky'. In 2000 his program was ranked 8th in the Computer Shogi Championships held in Tokyo. Since 2001 he has been a system architect in Ekahau Inc., developing positioning systems for wireless networks. His current research interests are in probabilistic positioning algorithms and computer shogi.

Juha Sievänen has studied Computer Science at University of Helsinki, where he also worked several years as a system administrator and a researcher in a High Performance Gigabit I2O Networking Software (HPGIN) research group specializing in Linux device drivers. After this he worked at the Nokia Research Center in Helsinki before joining Ekahau, Inc. in January 2001, where he currently works as a WLAN specialist.